DOCUMENT RESUME

ED 044 890 40 EC 031 000

AUTHOR	Pickett, James M.
TITLE	Research on Frequency Transposition for Hearing Aids. Interim Report.
INSTITUTION	Gallaudet Coll., Washington, D.C.
SPONS AGENCY	Bureau of Education for the Handicapped (DHEW/OE), Washington, D.C.
BUREAU NO	BR-182071
PUB DATE	Feb 70
GRANT	OEG-2-7-070070-1522
NOTE	20p.
EDRS PRICE	EDRS Price MF-$0.25 HC-$1.10
DESCRIPTORS	*Audio Equipment, Auditory Perception, *Aurally Handicapped, *Exceptional Child Research, *Hearing Aids, Sensory Aids, Sensory Training

ABSTRACT

To investigate impaired residual disciminiation for low-frequency formants and its influence on electronic compensation effectiveness, evaluations were made on impaired discrimination for speech formants, synthetic enhancement of consonants, wearable transposer aids, and a speech perception survey. Results showed that certain persons with severe sensorineural hearing loss have low frequency discimination for the frequency location of speech formants which is nearly normal, and that structured training is necessary if the formant frequency is above 250 Hz. The synthetically enhanced consonants appeared to be easier to discriminate than naturally spoken consonants. Due to electronic problems, no conclusions were reached concerning the wearable transposer aids. The speech perception survey indicated that low-frequency vowel sounds are more discriminable to the poorest sensorineural listeners than high frequency vowel sounds. (Author/RD)

ED044890

OE/BEH:EC
Project 182071
PA 40

INTERIM REPORT
Project No. 182071
Grant No. OEG-2-7-070070-1522

RESEARCH ON FREQUENCY TRANSPOSITION FOR HEARING AIDS

James M. Pickett, Ph. D.
Gallaudet College
Washington, D. C. 20002

February 1970

Department of Health, Education, and Welfare

U. S. Office of Education
Bureau of Education for the Handicapped

EC 031 000 E

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

INTERIM REPORT
Project No. 182071
Grant No. OEG-2-7-070070-1522

RESEARCH ON FREQUENCY TRANSPOSITION FOR HEARING AIDS

BACKGROUND

The purpose of this research is to develop and test new electronic methods for altering speech so as to better compensate for severe hearing losses of certain types. Current conventional hearing aids are adequate for overcoming conductive losses and some mild perceptive losses. However, we believe that new methods are needed to compensate, even partially, for the more severe "discrimination" losses associated with sensorineural forms of deafness.

Severe deafness of early onset is largely sensorineural and it causes grossly deficient speech communication. This deficit pervades all aspects of deaf education with the result that there is severe retardation of intellectual development. If better means can be found for alleviating the deficient speech communication of deaf persons, large improvements would occur in their education.

The number of persons in the United States who are severely handicapped in speech communication, because of deficient capacity for auditory discrimination, is on the order of 300,000. This type of deficiency is not presently remediable.

The vocational, educational, and social importance of speech communication is well known. Adequate speech reception for the persons mentioned above is not available through current hearing aids, nor through conceivable extensions and improvements of present design principles. Lip-reading and sign-language serve limited communication goals, but they can not approach the speed, flexibility, and scope of speech communication.

So far as we know at present, there are two ways to alter speech in an attempt to better fit the residual hearing of a hearing-impaired person. One method is to selectively amplify those speech frequencies which can not be heard due to the poor auditory sensitivity. We will call this method frequency-amplitude compensation. The other method is to radically change the frequency patterns of speech so as to make better use of hearing regions where sensitivity may remain less affected. We will call this method frequency transposition. Both methods have advantages and disadvantages, some of which may be complementary for improving speech reception, depending on the type of hearing loss, and particularly on the characteristics of the residual discrimination capacity.

CURRENT STATUS

The current status of this problem is briefly reviewed below under three headings: (1) Nature of hearing loss for speech, (2) Frequency-amplitude compensation for hearing loss, and (3) Frequency transposition for hearing aids.

1. Nature of Hearing Loss for Speech. The acoustic information of speech consists of a time-flow of fluctuating amplitude-frequency patterns. Persons with impaired hearing often suffer their worst losses in frequency regions that are the most important in speech. According to the classical studies of Fletcher (1953), the most dense region, containing 50% of the information, ranges from about 1000 to 3200 Hz, which unfortunately is a range where persons with sensorineural hearing impairment may receive little or no information. A person with usable hearing only below 1000 Hz receives only 25% of normal speech information; below 500 Hz only 10% of the information would be received. More recent research on the essential details of speech frequency patterns (Liberman, 1957) has tended to confirm that normally only a few of the distinctive speech cues lie below 500 Hz.

It might be hoped that hearing-impaired persons could make more efficient use of the low-frequency speech cues. However, some results from a study by La Benz (1956) strongly imply that this is not the case. Normal listeners, and 22 listeners with sensorineural hearing losses, averaging 40 dB, received speech from which most of the energy above 500 Hz had been removed by filtering. The average reception performance of the sensorineural cases was the same as that of the normals (30% PB words correct). Comparisons, in our own laboratory, have indicated that discrimination of frequency spectrum differences at 250 Hz, by listeners with moderate to severe cochlear impairment, is about as good as for normal listeners (see results of this work below). Thus we suspect that the basic reason for poor speech reception through low-frequency residual hearing is simply that there are few consistent cues to the basic speech distinctions that are normally present in the low-frequency speech energy.

2. Frequency-Amplitude Compensation for Hearing Loss. Sensorineural hearing losses, as characterized by standard audiometric measures, generally show progressively declining sensitivity as one proceeds from lower to higher frequencies. In addition, the range of loudness between the raised threshold and the pain threshold is usually considerably narrowed. For these reasons it would seem that hearing aid performance for these cases would be improved by emphasizing the amplitudes of the middle and high frequencies of speech and by limiting the range of amplitude fluctuations so that the weak sounds are above threshold but the strong sounds are below the pain level. Former tests of frequency emphasis and amplitude limiting for sensorineural losses have not been extensive. The results are summarized in the following sections for these studies where experimental data have been presented.

Frequency-response Emphasis. The most extensive study of design criteria for hearing aids remains the one carried out at Harvard University by Davis, et al, (1947). Careful measurements were made of speech reception through various combinations of tilted frequency-response, different amounts of amplification, and amplitude limiting. There were 15 hearing-impaired subjects, most of whom had only very mild discrimination losses for wide-band aided speech. Five sensorineural subjects (designated MC, JH, PP, IS, and WW) had more serious losses; their audiograms sloped downward about 5 to 10 dB/oct.

Eight ears of the sensorineural subjects were tested using different conditions of frequency-response emphasis as follows: a low-pass response

sloping downward about 9 dB/oct., a "flat" response which sloped downward about 3 dB/oct., a high-frequency emphasis sloping upward about 3 dB/oct., and another high-frequency emphasis sloped upward about 9 dB/oct.

In the test results the "flat" and high-frequency circuits gave maximum PB-word reception ranging from 62 to 92% correct. The high-frequency responses were better than "flat" for 5 of the ears, equal for 2, and worse for one.

The extreme high-frequency emphasis was not noticeably superior and so we might conclude, as the authors did, that any attempt to "mirror" a patient's audiogram in his hearing aid would not lead to improved reception.

Shore, Bilger, and Hirsh (1960) made systematic measures of speech reception by five sensorineural patients through each of four different hearing aids as set on two different "tone" settings which changed the frequency response. In some cases the two response curves under comparison were a rather flat response vs an added high-frequency emphasis obtained by increasing the contribution of sharp response peaks in the region 2000 to 3500 Hz. This condition was associated with decreases in reception. There were two cases (Patient 4, Aids B & C) where attenuation of peaked response around 3000 Hz produced large increases in reception, 12 to 17% points in percent correct PB words. This occurred even though the total frequency band passed by the aid was substantially reduced by the attenuation. Conductive patients in the same study did not show any consistent susceptability to sharp response peaks. It is possible, then, that the transient "ringing" effects that are introduced in the reproduced sound by sharp response peaks are more detrimental to sensorineural cases than to those with purely conductive losses.

Amplitude Limiting. In the Harvard study, progressive peak clipping at high gain levels was applied to the speech signal. The clipping level was normally 124 dB. When this was lowered to 112 dB for one sensorineural subject, thus reducing the peak levels received, the high-frequency emphasis became relatively more efficient than before. A later series of tests with this subject compared reception under amplitude compression by automatic gain control with reception under peak clipping which simply limits the peaks abruptly. Compression was found to be slightly better than peak clipping at high sound levels.

More recent studies of amplitude compression for hearing-impaired listeners have been somewhat equivocal. Parker (1953) found that compression considerably improved reception for some of his cases of "inner ear" deafness, but not for others. However, Caraway (1964) found that when a constant peak power is employed, amplitude-compressed speech was only slightly, if at all, more intelligible to sensorineural cases. Both Parker and Caraway used very rapid release times (about 1 msec.) for the automatic gain changes in their compressors, thereby enabling more amplification of the weaker speech sounds when they immediately followed strong sounds. Lynn and Carhart (1963), in a systematic investigation of onset and release times, found that rather long release times (150 msec. or more) produced larger advantages for compressed speech relative to uncompressed speech in cases similar to those of Caraway. In the Harvard compressor a release time of 200 msec. was used.

A rapid onset and release time for automatic gain control will produce distortion of waveshape, particularly when low-frequency vowel energy is dominant.

None of the recent studies have employed frequency response alterations in connection with amplitude limiting, as was done in the Harvard study. When we consider that the vowel energy in the region of the second formant (800 to 2500 Hz) is often quite low in amplitude, relative to the first formant, yet quite important for vowel and consonant identification, (Liberman, 1957), it would seem that emphasis of selected frequency regions should be employed before amplitude compression. Otherwise the stronger frequency components of a sound would cause the compressor to reduce the gain so much that important weaker components are below threshold. Caraway used a compressor having three frequency channels but she did not test frequency selective compression as a variable.

3. Frequency Transposition for Hearing Aids. It is possible, by the use of various techniques of speech analysis and synthesis, to transpose the speech information in middle- and high-frequency ranges down to lower ranges. Thus the more dense regions of speech information can be brought within the range of low-frequency residual hearing. The methods of transposition change the frequency range of the acoustic patterns but they do not change the overall time patterns. A number of authors have suggested that a hearing aid using this principle may provide improved speech communication especially for those with very severe hearing loss (Denes, 1964; Johansson, 1966; Oeken, 1963; Piminow, 1962; Raymond and Proud, 1962; Tiffany and Bennett,1961). There have been a number of controlled systematic tests of transposition for deaf subjects.

In one experiment (Oeken, 1963), the transposer operated by removing, during short intervals, half of the total speech time; then the remaining signal segments were joined and reproduced by playback from magnetic storage at half the original speed, thus restoring the original time patterns and dividing the original frequencies by a factor of two. It should be noted that this method of transposition may involve some distortion if the segments removed are not small. In Oeken's papers the size of segments was not specified.

Deaf subjects were trained on interpreting spoken words that were thus transposed; their identification of the words improved with practice but the training also improved their identification of normal, non-transposed words to an even higher level of performance than for the transposed words.

There was no provision for removing the intense low-frequency sound which must have resulted from the frequency-halving of the first-formant vowel components. For example the strong vowel components in the region 300 to 800 Hz were transposed to the region 150 to 400 Hz. The simultaneous weaker components in the region 800 to 2500 Hz were transposed to the range 400 to 1250 Hz. But they were not emphasized in any way. Therefore, they may not have been audible to those sensorineural subjects who have much more sensitivity in the lowest frequencies.

A transposer designed and tested by Johansson (1966) leaves the vowels

largely unchanged and transposes only the higher frequency energy to a frequency region below that of the vowels. The method was to separate the high-frequency speech components above 3500 Hz, mix them with a carrier of 5000 Hz, and then to transpose this signal to a region below 1500 Hz. Systematic tests with this transposer showed dramatic improvements in identification of fricative consonants by profoundly deaf children. After transposer training, identification using a conventional aid showed no improvement over the previous low performance.

It appeared, from the transposer results available in 1966, that a research program should be pursued to study frequency transposition in some detail. The program should investigate the results of transpositions in selected frequency regions. Also the listeners' patterns of discrimination loss as a function of frequency should be taken into account.

4. Rationale of Program. It seems likely that effective electronic compensation for losses in speech discrimination will depend on characteristics of residual discrimination capacities. In cases where there is usable capacity remaining in the range 800 to 2500 Hz, appropriate frequency-response emphasis, followed by amplitude compression, should improve reception of vowel information, and of vowel transitional cues adjacent to consonants. In cases where only low-frequency discrimination remains, downward transposition of the range 800 to 2500 Hz may be useful during the vowel phases of speech, whereas during consonant phases a different transposition should be used which transposes speech sounds from the range above 2500 Hz. Intermediate cases may require combinations of transposed and non-transposed signals.

Therefore, a program of research was begun to investigate impaired residual discrimination for low-frequency speech formants and discrimination of certain aspects of frequency pattern changes using exaggerated formants.

In addition, a transposer system was built for laboratory experimentation and pilot trials were begun on a wearable transposer hearing aid. A brief description of activities and results follows.

MEASUREMENTS OF IMPAIRED DISCRIMINATION FOR SPEECH FORMANTS

This section of the report is a condensation of a forthcoming article by Pickett and Martony (1970):

Ten subjects with moderate sensorineural hearing losses were selected and screened audiometrically. These subjects were tested for vowel formant discrimination at low and middle frequency ranges. A second group of six subjects with more profound losses was also selected and tested with low-frequency vowel formants. The threshold hearing levels of the groups are shown in Fig. 1, where the "sloping" group is more profoundly deaf in the important speech range above 1000 Hz. Normal subjects were also tested.

The procedure for testing was as follows. The sounds to be discriminated were generated by a vowel synthesizer with electronically tunable formant resonators. The output of the synthesizer was a vowel-like wave.

The discrimination tests were controlled by a programming system which presented the sounds, shifted the formant frequency appropriately, and

then received the listener's responses. Three sounds were presented on each trial, one of which was selected at random to have a higher formant frequency relative to the formant frequency of the other two sounds which were identical. The formant frequency of the two identical sounds is called the reference frequency, F. On each trial the program set the amount of difference, ΔF, between the reference and the higher formant; the amount of difference depended on the listener's success on the preceding trial in identifying which of the three sounds was different from the other two. If the listener was correct, the difference was made smaller on the next trial; if he was wrong the difference was made larger for the next trial. In this manner a run of trials proceeded toward a level of difference near the listener's discrimination threshold and then more or less oscillated above and below the threshold. Usually a minimum of 15 trials were necessary to establish an oscillating pattern over the last 5 to 10 trials; occasionally as many as 30 trials were necessary; typically 20-25 trials were made for each run to threshold. The experimenter concurrently recorded the series of differences presented on a run and decided when to terminate the run; then he estimated the threshold by examining the pattern of differences presented over the final portion of the run.

A run to threshold always began with a very large difference. The listener responded on a set of three push-buttons. After each response by the listener, his choice was stored, by lighting a light under the response button, for his comparison with the correct answer which was indicated by flashing the light under the correct button after he had pushed a final posting button. Until the post button was pushed, the listener was free to change his response.

The procedure of beginning each run with a very large difference and providing immediate knowledge of results was adopted in order to produce rapid learning toward the maximum possible auditory discrimination, i.e. the lowest threshold. In addition, careful instructions were repeatedly presented in the early stages of the series of test sessions and care was taken to express approval to the listener when his performance improved; also, whenever necessary as a training procedure, trials were presented at a large difference level, informing the listener as to which of the three sounds would be different on each trial. Despite these attempts to promote rapid learning, improvements in discrimination continued over an extended period of time under some conditions, as will be seen below in the results.

For the single formant tests, discrimination runs to threshold were made at four frequency positions of the reference formant, F = 205, 275, 400, and 825 Hz. These frequency conditions were selected for testing in more or less random order by the experimenter.

Test sessions lasted about 50 minutes with a 5-minute break midway and a few shorter breaks between test conditions; usually 6 to 8 runs to threshold were made per session. Sessions were scheduled on two separate days each school week; testing began in July, recessed at the end of August and resumed for November through March except for a two-week Christmas recess.

The measure of discrimination used for summarizing the results is the mean $\Delta F/F$ based on estimated threshold from two consecutive runs. When $\Delta F/F$ is large, discrimination is poor; when it is small, discrimination is

good. The hearing-impaired listeners were divided into two groups according to amount of hearing loss at the audiometric frequency nearest the reference F (The mean hearing losses, H.L., are given on the figure). Group mean threshold $\Delta F/F$'s are shown in Fig. 2 for the four frequency positions of F and as a function of the cumulative number of runs at each F.

We note first in Fig. 2 that, except at F = 205, the group of normal listeners had better initial discrimination and more rapid learning to their best level than did any of the hearing-impaired groups. It is also apparent that the impaired groups, at their best, are as good as the normal group at the low positions of F, 205 and 400 Hz, but that the groups with the more severe losses (86.5 and 98.7 dB) at 500 and 1000 Hz do not reach the low discrimination threshold of the normal group at F = 400 and F = 825 Hz.

There is a rough correlation at F = 400 and 825 Hz between hearing loss and size of threshold at the best level of discrimination. The mean threshold for the larger loss group ranges about 2 to 4 times that of the smaller loss group.

A brief set of tests were carried out with the impaired subjects to relate synthetic vowel discrimination to the discrimination of natural vowels spoken in words. The mid- and back-vowel series was chosen for testing since the frequency locations of formants in these vowels cover a frequency range similar to the range of the formant in our synthetic vowels. The vowels used were / a, ʌ, ɔ, o, ʊ, u/; each of these six vowels was paired five times with each of the other five vowels to make a total of 150 two-choice test items.

The subjects scored an average of 87% correct in the two-choice spoken vowel test. The subjects were ranked according to vowel discrimination errors and according to size of $\Delta F/F$ threshold. The rank of each subject according to number of errors was the same as his rank according to size of his mean $\Delta F/F$ threshold for synthetic vowel discrimination at F = 275, 400, and 825 Hz. We concluded that synthetic vowel discrimination may be a useful predictor of natural vowel discrimination.

A set of control tests were carried out to compare $\Delta F/F$ discrimination by auditory listening and by tactual discrimination of the same sound.

The procedure and apparatus for the measurements of tactual discrimination were the same as for auditory discrimination. The tactual runs to threshold were made with the earphone cushion placed on the palm of the hand and held there tightly by the subject, or alternatively, wearing the headset with earphone located on the cheek in front of the test ear. The levels for tactual runs were the same as for auditory runs. Careful instruction and practice was given, encouraging the subject to feel for any vibratory change as a basis for discrimination, such as changes in "smoothness" or in intensity. The tactual runs were made only at the lower F locations and they were interspersed with auditory runs at the same frequency locations and sound pressure levels.

It was concluded from the tactual controls that those ΔF thresholds above about 0.10 may represent vibratory discrimination or a combination of

vibratory and auditory discrimination, and probably based on sensations of vibratory intensity. Thus for the profoundly impaired listener, at high sound levels, his formant discrimination at low frequencies may not be distinguishable from tactual discrimination. When the F-discrimination is at a level smaller than $\Delta F/F = 0.10$ or does not require sound levels higher than about 100 dB we believe that the discrimination is auditory. However, some more extensive control tests should be made on this point.

We now consider the relation between these results and the use of frequency-shifting and transposing for hearing aids. First of all, we would emphasize the long learning effects which occurred in our discrimination tests even with close juxtaposition of the sound changes and immediate knowledge of the correct answer. In the "casual" use of a transposing hearing aid, where there is no control over the sounds which occur, and knowledge of the result of a discrimination would often be equivocal, we would expect even longer learning periods. Therefore, the transposer training situation should be carefully designed to isolate the sounds to be discriminated. Even then some listeners may require considerable training to reach only fair discrimination above about 500 Hz.

If low-frequency shifted or transposed speech sounds differ from each other in frequency by less than about 5%, discrimination performance for these sounds would be close to threshold for deaf listeners. All the speech sound differences that depend on frequency discrimination of normal formant positions are larger than 5%. A few are between 5 and 15%, e.g., some of the vowels that are adjacent in the vowel triangle, such as /æ, ɛ/ and /I/. Most of the differences, however would be between 25 and 50%, and these should be discriminable by impaired listeners like those in the present study if they are proportionately transposed or shifted into a range below 1000 Hz and if the frequency spectrum is subjected to a moderate upward tilt. The above statements apply only to discrimination of the long, relatively steady-state sounds of speech which are about 150 to 300 msec. in duration. There are short and transient differences among speech sounds that are important in their discrimination. These results should not be applied to predictions of the discrimination of the transient speech sounds.

In measuring discrimination for the group of moderately impaired subjects, a vowel formant discrimination task was used where a low formant remained fixed and a second formant was added in the middle frequency range (1000 to 1600 Hz); the second formant was then varied in frequency to determine the subject's threshold using the same procedure as described above. These tests were carried out for different frequency spacing relations between the two formants. Some of the group of moderately impaired listeners were nearly as good as normal listeners in their formant discrimination, in that their discrimination thresholds were small (about 1%) and not affected by the formant spacing; others, however, appeared to be grossly inferior to normals and much worse when the variable formant was close to the low fixed formant.

It is felt that these discrimination defects will have an important bearing on the success of transposer hearing aids because some transposing

methods inherently reduce the formant spacing. Therefore it was decided to investigate these effects together with live-speech tests with transposer systems.

TESTS OF SYNTHETIC ENHANCEMENT OF CONSONANT CUES

The consonant sounds of speech are generally of a transient, dynamic nature in contrast to the steady vowel patterns used in our formant discrimination tests. However, we arranged to make special consonant sounds for tests to explore the possibility of electronic enhancement of consonant patterns for hearing-impaired subjects. Highly flexible methods for altering speech signals were made available to us through cooperation with Haskins Laboratories, a basic speech research laboratory at Yale University, New Haven, Connecticut. Haskins staff expressed a deep interest in our project and proposed that their system for speech synthesis be used to generate speech sounds for our tests that could be altered in frequency and in numerous other ways. The synthesis system is computer-controlled, a fact which makes it easy to generate alterations of speech in the time domain. For our preliminary tests we made a set of special consonant-vowel syllables that were stretched in time and exaggerated in formant structure; another set was made that was lowered in frequency.

The syllables were ka, ta, pa, ga, da, and ba. The formant circuits of the synthesizer were set at relatively narrow bandwidths with steep slopes on either side so that the formant peaks would be more prominent than in natural speech.

Four different versions of the set were synthesized, one with normal timing (X1), one with time stretched by a factor of two (X2), another with time stretched X3, and one time-compressed to 3/5 of normal time during synthesis, but played back at 3/5 speed for testing so as to lower all the frequencies by a factor of 3/5 but restore the normal timing. Natural utterances of the same set of syllables were also prepared.

Acoustic analysis of the syllables showed that the synthetic syllables had formant peaks that were better defined than those of the natural syllables. In addition the synthetic second formant amplitude was equal to that of the first formant, whereas, in the natural speech the second formant was about 5 dB lower than the first formant. Also the formant transitions of the synthetic syllables were longer and better defined than in the natural syllables.

Tests of identification of the syllables were carried out. Seven subjects with severe to profound deafness were tested. In a first set of preliminary tests, the syllables were used in pairs. Nine pairs were used, in a block of tests, testing the voicing distinction for each place category: labial, pa-ba, alveolar, ta-da, and velar, ka-ga; and also testing the place distinctions, labial-alveolar, alveolar-velar, and labial-velar, both for voiced and unvoiced consonants. Each test was repeated twice at separate times in the series of tests. For half the listening sessions, the series of tests began with the block of nine tests with the X2 syllables, then proceeded to X1, X3, and Natural; for the other half of the sessions, the order was X1, X2, Natural, and X3.

During each test the results were tallied in a 2 x 2 stimulus-response matrix, with the subject making his identification response by pointing to one of the two response columns, which were labelled with the appropriate consonants; then the experimenter indicated the correct answer to the subject, by pointing to the correct stimulus row, and tallying the response in the proper cell. Thus the subject had feedback as to the correct answer.

The results were expressed in terms of percent correct response. Pooling all results under each condition, the best performance was obtained with time-stretching of X2, 77.7% correct, and with X3, 76.8% correct, both significantly better than Natural, 71.5%, and X1, 73.1%. Natural and X1 were not significantly different.

We then examined individual listener performance on the three different classes of consonant distinctions in our test pairs, i.e. distinctions of voicing (/pa-ba/, /ka-ga/, and /ta-da/), of voiced place (/ba-da/, /da-ga/, and /ba-ga/), and of unvoiced place (/pa-ta/,/ta-ka/, and /pa-ka/). The listeners were divided into two groups, one group of three subjects with the best performance, and the other group consisting of the remaining four subjects, based on individual ranks in performance on all six tests of place distinctions pooled. The results of this analysis are shown in Fig. 3. A vertical distance module appears in each panel of the figure which gives the size of any difference between measures that must be exceeded to be statistically significant.

The voicing distinction was heard better than the place distinctions for both groups of listeners. Also the better group heard both voicing and place significantly better than the poor group except for X3 voicing and Natural unvoiced place.

Comparing Natural and X1 synthetic, the synthetic was significantly superior for voicing, significantly inferior for voiced place distinctions by the poor group, and significantly better for unvoiced place distinctions by the better group. In the two remaining comparisons, Natural and synthetic X1 did not differ.

The stretching of the synthetic syllables from X1 to X3 significantly improved the voicing distinction for the poor group but significantly interfered with this distinction in the better group. With voiced place, the stretching had no effect in the better group but it improved the distinction for the poor group; for this case the drop in performance from X2 to X3 was not significant. With unvoiced place distinctions the only significant effect of stretching was the drop from X2 to X3 for the better group of listeners.

For the better group of listeners, the distinction alveolar <u>vs</u> velar was significantly more difficult than labial <u>vs</u> alveolar and labial <u>vs</u> velar; for the poor listeners this tendency was present but not significant statistically.

Up to this point in our experiments on identification of synthetic syllables, only a single pair of syllables was used in any one block of

training-testing trials. Further preliminary experiments were made with larger sets of the synthetic syllables as sets for training and testing identification. These experiments were carried out using an automatic speech-testing system which was constructed during the last half of the report period (see description of this system below).

Generally all six syllables were used, unless performance was nearly at chance level. Then trials were run with subsets of only three syllables as training for subsequent tests with all six syllables. Three syllable conditions were tested, Natural, Synth X2, and synthetic lowered in frequency by 3/5 (Synth X3/5). The same subjects were used as in the tests with pairs of syllables. Only about eight tests have been run thus far on each condition on each subject. Preliminary estimates of performance in these tests are consistent with the previous results. For Nat and Synth X2, performance as a whole is about half way between chance and perfect performance with Synth X2 slightly superior to the Natural syllables.

However, with the Synth X3/5 (40% lower in all frequencies but with normal timing patterns) performance was poor for the poorer subjects and about the same as Synth X2 for the better subjects. The Synth X3/5 syllables sound quite different from natural speech and thus present a serious re-learning problem. Our results are still only preliminary, but it is extremely important to determine what re-learning problems will be encountered with frequency-divided speech and whether they can be overcome. For example, if they cannot be overcome by the adult deaf listener, it will be necessary to carry out frequency-dividing studies only with very young children.

DESIGN AND CONSTRUCTION OF AUTOMATIC SPEECH-TESTING SYSTEM

An automatic speech-testing system was constructed to facilitate the training and testing of speech sound identification. This system consists of an eight-channel precision tape recorder, and associated logic circuits for selecting sounds recorded on the tape, presenting them to a listener, receiving his identification responses, and displaying the test results in the form of an accumulated matrix of frequency counts of all the responses, by stimulus categories and by response categories. The subject responds to each stimulus sound by pushing a labelled response button after which the system informs the subject of the identity of the sound by flashing a light under the correct button. The stimulus-response matrix of results allows us to see the structure of perceptual similarities between stimulus sounds as well as to assess overall success in identification. The automatic testing system has been in use for about one year in tests described in this report.

Another feature of the automatic speech-testing system is an infinitely variable playback speed. This enables us to use the playback as a frequency-dividing device, simply by adjusting the playback speed to be slower than that of the original tape recording. Of course slower playback stretches the duration of the sound played back, but this can be normalized by time-compressing the original sound before recording it. For example, for one

of our test conditions described above, with synthetic syllables which were frequency-divided by a factor of 3/5, the original syllables were time-compressed by 3/5 as they were generated by the Haskins Laboratories synthesizer. This synthesis was made with normal frequency structure. However, for our tests, the playback speed was adjusted to 3/5 of the original recording speed, thus dividing the frequencies by this factor and expanding the compressed time patterns back to normal time.

CONSTRUCTION OF LABORATORY TRANSPOSING SYSTEM AND PRELIMINARY RESULTS

A flexible transposing system was constructed for transposing various frequency bands of speech either by the Johansson method or by a frequency-dividing method. The speech signal to be processed is first applied for analysis to a bank of 13 contiguous band-pass filters which separate the frequencies of the input speech for independent transposing. The frequency limits of the filters are given in Table 1.

TABLE I

Filter Number	Frequency Range
1	0-250 Hz
2	250-350
3	290-400
4	400-520
5	500-650
6	650-800
7	780-1000
8	1000-1300
9	1300-1680
10	1600-2000
11	2000-2500
12	2500-3200
13	4000-6200

Any set of filter outputs can be combined and fed to any one of three frequency dividers and two heterodyne transposers, and then the resulting signals are mixed together to form the final output signal.

For preliminary tests the system was arranged with a heterodyne transposition of bands 8-10, that is 1000-2000 Hz, covering most of the second formant range of male talkers, and of band 13, 4000-6200. The second formant range is transposed to a range 200 to 1200 Hz and the frequency scale is inverted, so that components near 2000 Hz are transposed to about 200 Hz and the components near 1000 Hz are transposed to 1200 Hz. The range 4000-6200 Hz was transposed to the range 700 Hz (representing 4000 Hz) to 1500 Hz. This high frequency transposing circuit is similar to that of Johansson and is intended to make audible, in a low-frequency range, certain consonant components normally in the high-frequency range.

Preliminary tests were carried out with vowel sounds. A set of eight different vowels were spoken into the system and their transposed

versions were recorded on the automatic speech-testing system. Then both the normal and transposed vowels were played back for identification tests. Six subjects have been partially tested thus far. One subject, with rather poor performance, could identify the transposed vowels slightly better than the normal vowels. It appears that considerable practice and training will be necessary for subjects of this type. The better subjects learned rapidly to identify the transposed vowels to a fairly high level of success; however, these subjects could also identify the normal vowels with the same success.

EVALUATION OF WEARABLE TRANSPOSER AIDS

The plan for this part of the project is to carry out field trials with two types of transposer hearing aid. One is a body-worn aid, of the Johansson type (see pp. 4-5 above), the Model Tp. 64, manufactured by Oticon. The other transposer aid is a binaural ear-level aid in prototype development by the Acousis Company. So far, we have not been able to start trials with these aids due to technical problems that have arisen in development or use.

We consider the Acousis aid to be particularly interesting because it is a high-gain instrument (70 dB) that uses ear-level microphones. This should enable many wearers to receive binaural cues to the location of speech sources and noise sources and thereby improve their aided speech reception.

SPEECH PERCEPTION SURVEY

A speech perception survey has been made to provide baseline data on perception of normal vowels. One hundred deaf students at Gallaudet College were tested using a multiple choice test form in order to minimize effects of language factors. During the test a subject heard a series of stimulus words delivered at a sound level that was adjusted to his hearing loss. After hearing each word, the subject tried to identify the word as one of six words printed as response choices; the response choices differed only in the vowel. For example, for the stimulus word beet, the six response words were beet, bait, bet, boot, boat, and bought.

The results were analyzed separately in four subgroups of 25 students representing four levels of overall performance on the test. The results showed that the average vowel reception, respectively for the four subgroups, was 37, 41, 76, and 88% correct. In the two poorer subgroups, vowels with low-frequency distinctions (or cues to identification) were perceived better than vowels with distinctive cues at higher frequencies. The difference was 13% points for one subgroup and 14 points for the other subgroup. The two better subgroups perceived vowels somewhat better generally, but with no difference between low vowels and high vowels.

Facts like these should enable us to better judge the basic discrimination capacity for vowel cues as a function of the amount of impairment. We can also get a rough idea of how much improvement might be gained by transposing high-frequency vowel cues to a lower region. Also we note that

vowel discrimination is severely impaired for the poorer half of our sample population.

CONCLUSIONS

Tests of impaired discrimination for speech formants. Some persons with severe sensorineural hearing loss have low-frequency discrimination for the frequency location of speech formants that is nearly normal; if the formant frequency is above about 250 Hz, the listener with sensorineural impairment may need a large amount of training before attaining normal discrimination; this training should be specifically structured and well-organized.

Tests of synthetic enhancement. The synthetically enhanced consonants were somewhat easier to discriminate than naturally spoken consonants. Synthetic syllables with lowered frequency structure did not seem more discriminable for sensorineural listeners than the same syllables in a normal frequency range. Time-stretching of syllables (with no frequency shift) gave small improvements in discrimination.

Evaluation of wearable transposer aids. No conclusions were reached because evaluation tests have not got started, due to electronic problems.

Speech perception survey. Low-frequency vowel sounds (in natural speech) are more discriminable to the poorest sensorineural listeners than high-frequency vowel sounds. For the better impaired listeners there was about equal discrimination.

RECOMMENDATIONS

1. It is necessary to obtain further knowledge of impaired sensorineural discrimination, as a basis for developing special speech-coding hearing aids.

2. Auditory training methods should be developed that employ sound stimuli that is carefully controlled, probably through the use of synthetic speech patterns.

3. A long coordinated program of research should be carried out on recommendations 1 and 2. The fundamental problems are very complex and our knowledge of impaired sound discrimination is very meager.

REFERENCES

Caraway, B. J. Effects on speech intelligibility of dynamic range reduction with a compressor amplifier. Ph. D. Thesis, Northwestern Univ., 1964.

Davis, H., Stevens, S. S., Nichols, R. H., Hudgins, C. V., Marquis, R. J., Peterson, G. E., and Ross, D. A. Hearing Aids: An Experimental Study of Design Objectives. Harvard Univ. Press, Cambridge, Mass., 1947.

Denes, P. On the motor theory of speech perception. Paper presented at the Symposium on Models for the Perception of Speech and Visual Form, A. F. Cambridge Research Labs, Boston, Mass., 11-14 November 1964.

Fletcher, H. Speech and Hearing in Communication. Van Nostrand, New York, 1953.

Johansson, B. The use of the transposer for the management of the deaf child. Int. Audiol., 5, 1966, 362-372.

LaBenz, P. J. Potentialities of auditory perception for various levels of hearing loss. Volta Bureau Reprint Number 683, 1956.

Liberman, A. M. Some results of research on speech perception. J. Acoust. Soc. Amer., 29, 1957, 117-123.

Lynn, G., and Carhart, R. Influence of attack and release in compression amplification on understanding of speech by hypoacusics. J. Speech Hear. Dis., 28, 1963, 124-140.

Oeken, F. W. Frequenztransposition zur Horverbesserung bei Innerohr-schwerhorigkeit. Archiv Ohren usw. Heilk. u. Z. Hals usw. Heilk., 181, 1963, 418-425.

Parker, C. D. The effects of the reduction of "short time fatigue" on speech intelligibility for "perceptively" deafened individuals. Ph. D. Thesis, State Univ. of Iowa, 1953.

Pickett, J. M., and Martony, J. Low-frequency vowel formant discrimination in hearing-impaired listeners. J. Speech Hear. Res., 1970, in press.

Piminow, L. L'application de la parole synthetique dans la correction auditive. Acustica, 12, 1962, 285-290.

Raymond, T. H., and Proud, G. O. Audiofrequency conversion. Arch. Otolaryngol., 76, 1962, 436-446.

Shore, I., Bilger, R. C., and Hirsh, I. J. Hearing aid evaluation: reliability of repeated measurements. J. Speech Hear. Dis., 25, 1960, 252-170.

Tiffany, W. R., and Bennett, D. N. Intelligibility of slow-played speech, J. Speech Hear. Res., 4, 1961, 248-258.

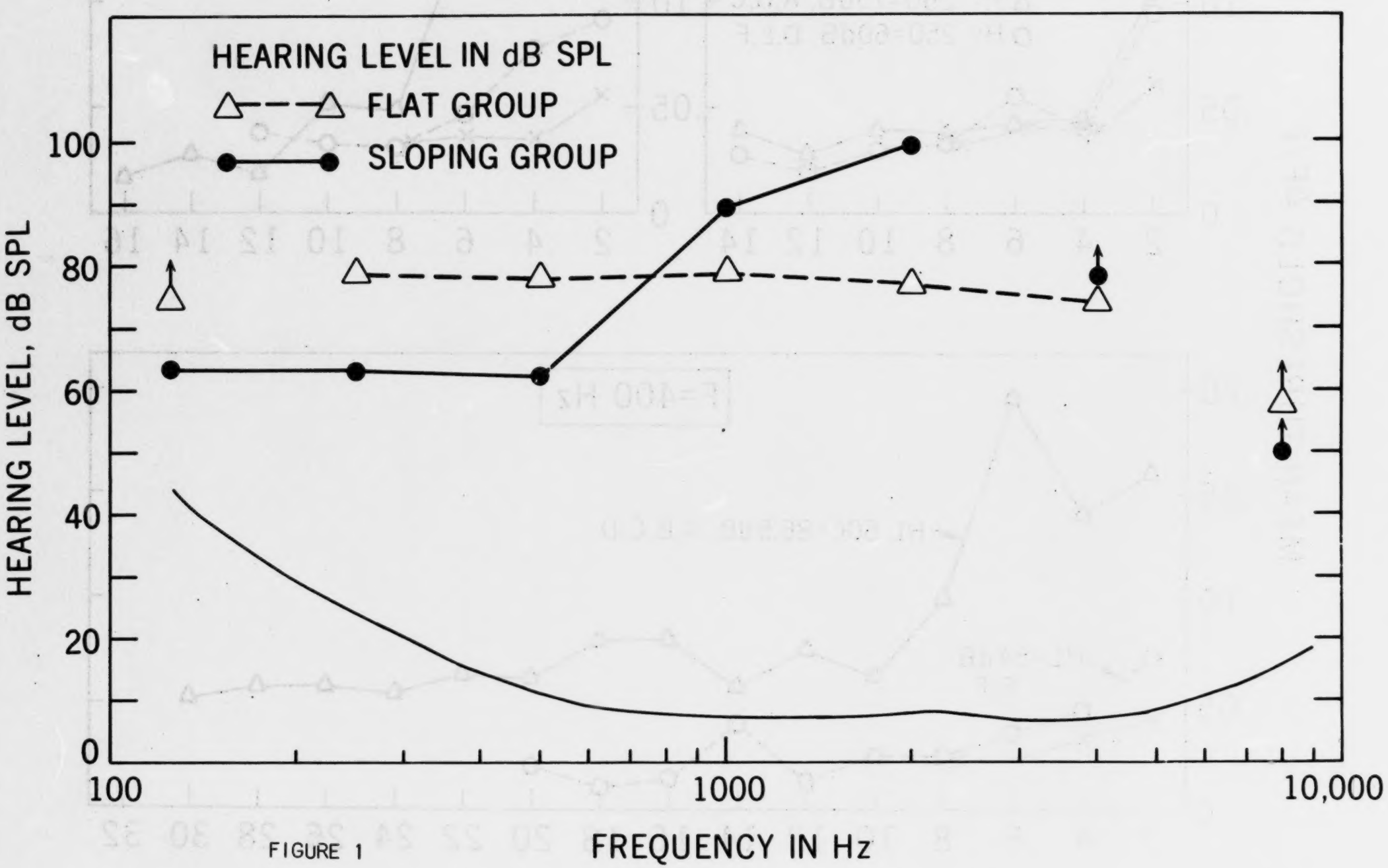
HEARING LEVEL IN dB SPL
FLAT GROUP
SLOPING GROUP
HEARING LEVEL, dB SPL
100
80
60
40
20
0
100
1000
10,000
FREQUENCY IN Hz

FIGURE 1

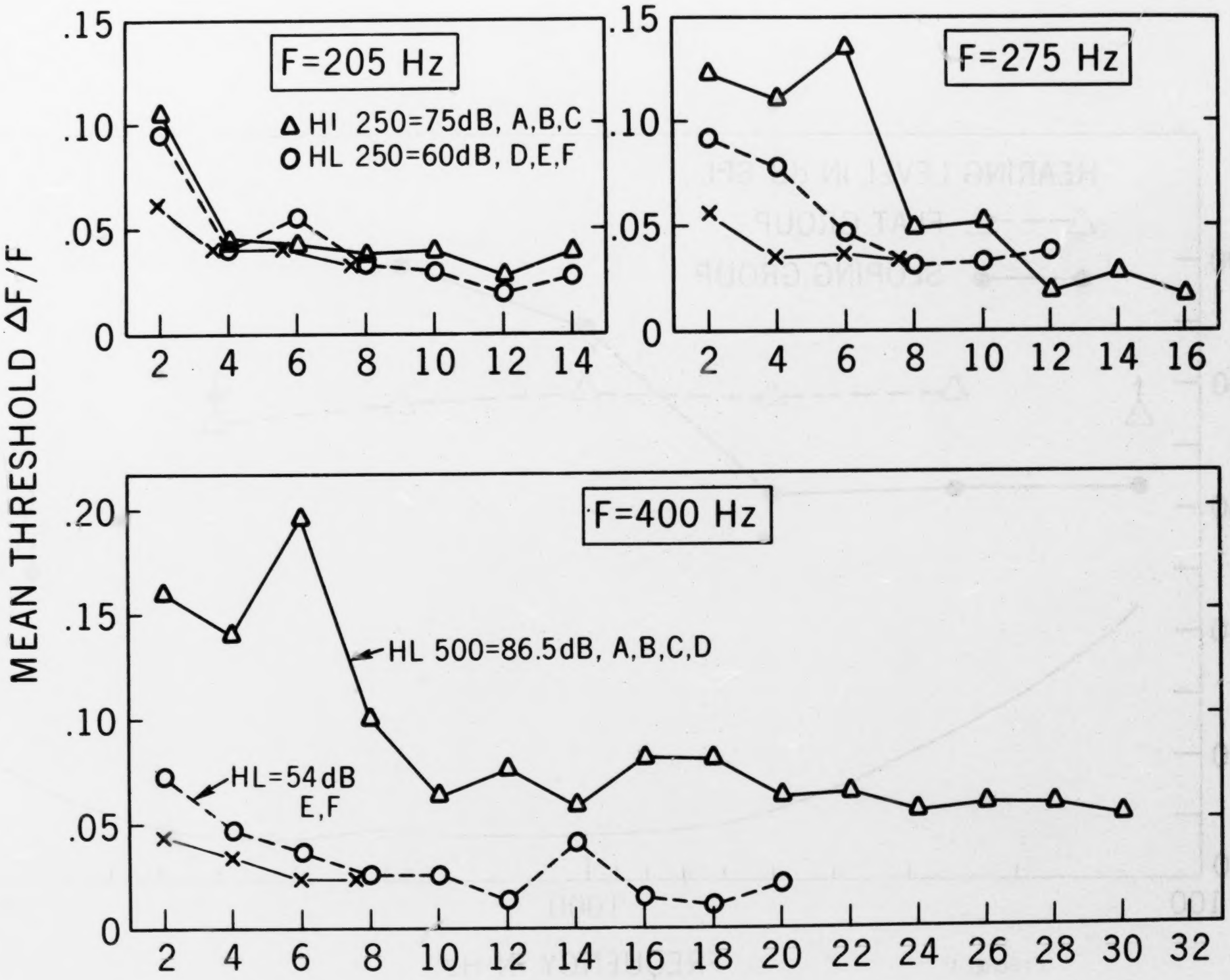

FIGURE 2, TOP HALF

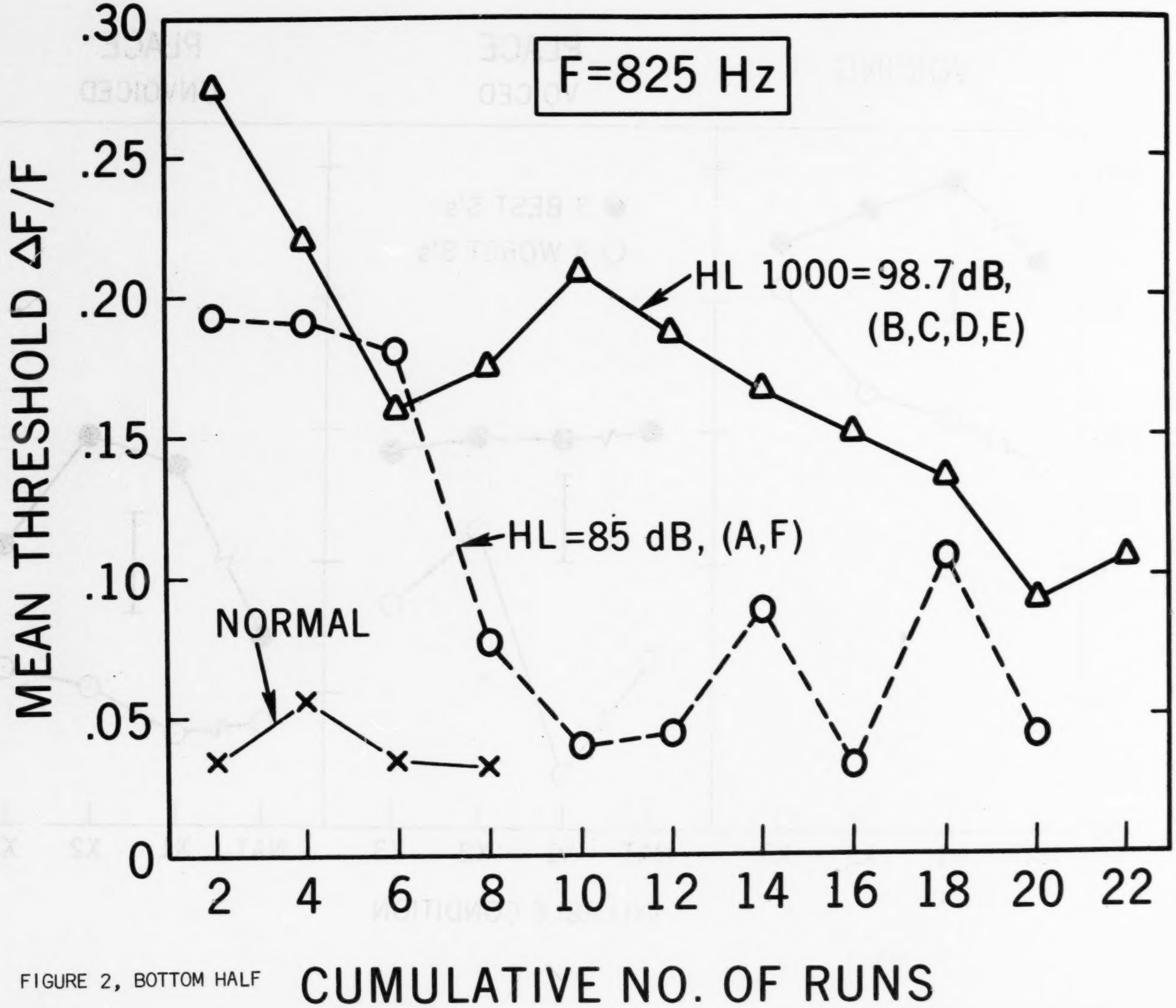

FIGURE 2, BOTTOM HALF

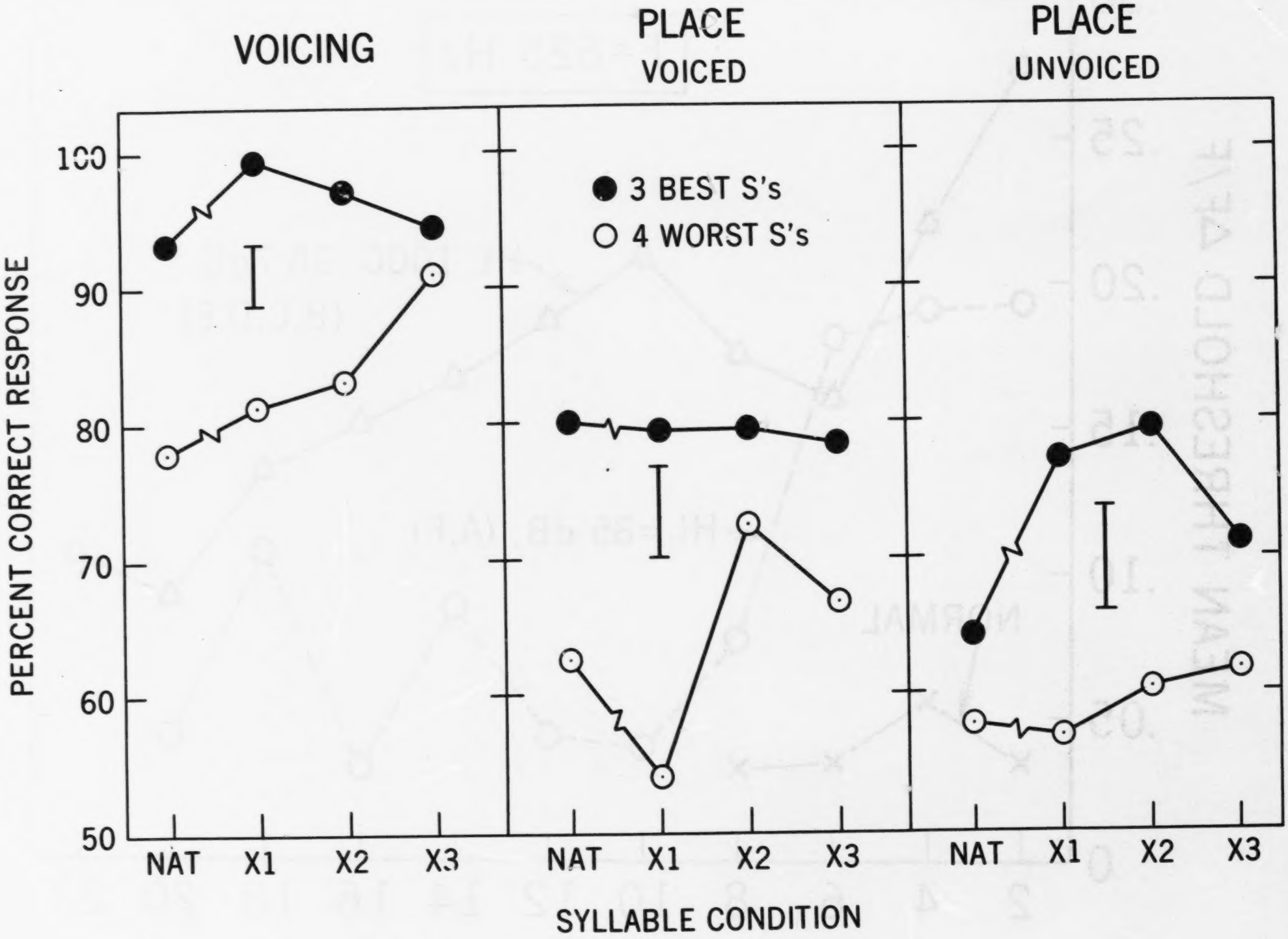
VOICING
PLACE
VOICED
PLACE
UNVOICED
3 BEST S's
4 WORST S's
PERCENT CORRECT RESPONSE
100
90
80
70
60
50
NAT
X1
X2
X3
SYLLABLE CONDITION

FIGURE 3